

10-8-2020

African Americans and European Americans exhibit distinct gene expression patterns across tissues and tumors that are associated with immunologic and infectious functions and environmental exposures

Urminder Singh

Iowa State University, usingh@iastate.edu

Kyle M. Hernandez

University of Chicago

Bruce J. Aronow

Cincinnati Children's Hospital Medical Center

Eve Syrkin Wurtele

Iowa State University, mash@iastate.edu

Follow this and additional works at: https://lib.dr.iastate.edu/gdcb_las_pubs



Part of the [Cell and Developmental Biology Commons](#), [Genetics and Genomics Commons](#), and the [Immunology and Infectious Disease Commons](#)

The complete bibliographic information for this item can be found at https://lib.dr.iastate.edu/gdcb_las_pubs/262. For information on how to cite this item, please visit <http://lib.dr.iastate.edu/howtocite.html>.

This Article is brought to you for free and open access by the Genetics, Development and Cell Biology at Iowa State University Digital Repository. It has been accepted for inclusion in Genetics, Development and Cell Biology Publications by an authorized administrator of Iowa State University Digital Repository. For more information, please contact digirep@iastate.edu.

African Americans and European Americans exhibit distinct gene expression patterns across tissues and tumors that are associated with immunologic and infectious functions and environmental exposures

Abstract

The COVID-19 pandemic has affected African American populations disproportionately with respect to prevalence, morbidity, and mortality. Because gene expression profiles represent combined genetic, socioenvironmental, and physiological effects, and could provide therapeutic biomarkers and environmental mitigation strategies, we undertook a large-scale assessment of differential gene expression between African Americans and European Americans. To do this, we mined RNA-Seq datasets from normal and diseased (tumor) conditions whose metadata could be used to evaluate differential patterns. We observed widespread differential expression of genes implicated in COVID-19 and integral to epithelial boundary function, inflammation, infection, and reactive oxygen stress. Notably, expression of the little-studied F8A2 gene is up to 40-fold greater in African Americans. F8A2, like F8A1, encodes HAP40 protein, which mediates early endosome movement. African American gene expression signatures reveal increased number or activity of esophageal glandular cells and lung ACE2-positive basal keratinocytes. These findings have potential to establish prognostic signatures, refine approaches to minimizing risk of severe infection, and improve precision treatment of COVID-19.

Disciplines

Cell and Developmental Biology | Genetics and Genomics | Immunology and Infectious Disease

Comments

This preprint is made available through Research Square, doi: [10.21203/rs.3.rs-88890/v1](https://doi.org/10.21203/rs.3.rs-88890/v1).

Creative Commons License



This work is licensed under a [Creative Commons Attribution 4.0 License](https://creativecommons.org/licenses/by/4.0/).

African Americans and European Americans exhibit distinct gene expression patterns across tissues and tumors that are associated with immunologic and infectious functions and environmental exposures

Urminder Singh^{1,2,3}, Kyle M. Hernandez^{4,5}, Bruce J. Aronow⁶, and Eve Syrkin Wurtele^{1,2,3*}

¹Bioinformatics and Computational Biology Program, Iowa State University, Ames, IA, 50011, USA

²Center for Metabolic Biology, Iowa State University, Ames, IA, 50011, USA

³Genetics Development and Cell Biology, Iowa State University, Ames, IA, 50011, USA

⁴Department of Medicine, University of Chicago, Chicago, IL, 60637

⁵Center for Translational Data Science, University of Chicago, Chicago, IL, 60637

⁶Biomedical Informatics, Cincinnati Children's Hospital Medical Center, Cincinnati, OH 45229

*mash@iastate.edu

ABSTRACT

The COVID-19 pandemic has affected African American populations disproportionately with respect to prevalence, morbidity, and mortality. Because gene expression profiles represent combined genetic, socioenvironmental, and physiological effects, and could provide therapeutic biomarkers and environmental mitigation strategies, we undertook a large-scale assessment of differential gene expression between African Americans and European Americans. To do this, we mined RNA-Seq datasets from normal and diseased (tumor) conditions whose metadata could be used to evaluate differential patterns. We observed widespread differential expression of genes implicated in COVID-19 and integral to epithelial boundary function, inflammation, infection, and reactive oxygen stress. Notably, expression of the little-studied F8A2 gene is up to 40-fold greater in African Americans. F8A2, like F8A1, encodes HAP40 protein, which mediates early endosome movement. African American gene expression signatures reveal increased number or activity of esophageal glandular cells and lung ACE2-positive basal keratinocytes. These findings have potential to establish prognostic signatures, refine approaches to minimizing risk of severe infection, and improve precision treatment of COVID-19.

Introduction

The COVID-19 pandemic has infected over 31 million people and killed over 970,000 worldwide as of September, 2020 (<https://coronavirus.jhu.edu/map.html>). Its causative agent, the novel SARS-CoV-2, is an enveloped single stranded RNA virus that infects tissues including epithelial cells in the upper respiratory tract, lung alveoli, GI tract, vasculature endothelium, renal tubules, central nervous system, and myocardium¹⁻⁶. The complex combinations and severities of symptoms caused by SARS-CoV-2 include fever, cough, fatigue, dyspnea, diarrhea, thrombosis, stroke, acute respiratory failure, renal failure, cardiac failure; in some individuals these may lead to long-term disability or death^{2,5,6}. Differing patterns of disease may result from direct cellular infection, secondary inflammatory repercussions, and circulating immune and necrotic complexes from distal sites of infection and response⁷⁻⁹. How these attributes confer risk of increased disease severity to individuals is not well understood^{4,8,10,11}. Identifying individuals most at-risk for severe COVID-19 infection, and determining the molecular and physiological basis for this risk, is critical to enable more informed public health decisions, and improving our identification and use of precision interventions.

COVID-19 cases and deaths are disproportionately higher among African Americans in the US¹⁰. This disparity is caused in part by complex combinations of socio-economic factors, including underlying comorbidities, air quality, population density, and health care access¹⁰. Heritable factors in the human host also influence COVID-19 symptoms¹²⁻¹⁶. To date, several genetic determinants of COVID-19 severity have been partially elucidated. Genetic variants of Angiotensin-Converting Enzyme2 (ACE2), a major human host receptor for the SARS-CoV-2 spike protein, may be linked to increased infection by COVID-19¹⁵. Human Leukocyte Antigen (HLA) gene alleles have been associated with susceptibility to diabetes and SARS-CoV-2¹⁴. A

COVID-19 association at locus 9q34.2 spans several genes related to COVID-19, including blood type¹³. The genetic propensity in southern European populations for mutations in the pyrin-encoding Mediterranean Fever gene (MEFV) has been proposed to lead to elevated levels of pro-inflammatory molecules, a cytokine storm, and greater severity of COVID-19¹⁶. Multiple GWAS associations based on ancestry are beginning to emerge (<https://grasp.nhlbi.nih.gov/Covid19GWASResults.aspx>)¹³.

Gene expression is a reflection of a cell's composition and its spatial and developmental context in an organism. Modifying factors that determine gene expression span genetics, and physiological, environmental, and socioeconomic influences. In this study we seek to investigate potential differential expression of genes and pathways that may impact the severity of COVID-19 infection in African Americans. Research with macrophage cell lines has identified ancestry-related differences in innate immune response to bacterial pathogens, with cell lines isolated from individuals with African ancestry more likely to exhibit stronger inflammatory responses¹⁷.

Here, we utilize publicly available datasets to explore gene expression differences between African American and European American individuals. We analyze "non-diseased" tissues from the Genotype Tissue Expression (GTEx, <https://gtexportal.org/home/>) project, and, as representative of highly perturbed systems, tumor samples from The Cancer Genome Atlas (TCGA, <https://www.cancer.gov/about-nci/organization/ccg/research/structural-genomics/tcga>). In addition, we seek to unravel the cellular origins of ancestry-associated differential gene expression through the use of Human Cell Atlas single-cell datasets from esophagus and lung tissues. Taken together, our analyses reveal that European Americans and African Americans have significant differences in expression of multiple genes across most sample-types. Pathways likely to impact the severity of COVID-19 are enriched in these DE genes. In esophagus and lung, two tissues critical to early SARS-CoV-2 infections, population-differential gene signatures implicate specific cell lineages that are likely to alter viral disease severity. Finally, we highlight the importance of evaluating population-related impacts on gene expression in the context of cellular and organismal health and resilience to disease in the combined light of socioeconomic/environmental and genetic factors.

Results

In order to identify genes differentially-expressed (DE) between African American and European Americans we constructed an aggregated dataset of 7,142 RNA-Seq samples encompassing nine non-diseased tissues from GTEx and eight cancers from TCGA^{18,19}. Race assignments are self-reported in the metadata; however, many of the individuals sampled identifying as a single race may be from an admixed population^{20,21}. We analyzed data and metadata using MetaOmGraph (MOG)¹⁸, software that supports interactive exploratory analysis of large data to identify and distinguish patterns across multiple dimensions (Table 1 and Supplementary Table S1).

Genes are DE between populations in a tissue- and tumor-specific manner

DE genes were identified for each sample-type, as well as for pooled TCGA and GTEx data (Table 1 and Supplementary Tables S2-S28). The number of samples affects the power of the DE test (Table 1). To test for potential confounding factors that might explain gene expression pattern differences, we scrutinized differences between African American and European Americans populations controlling for biologically-relevant factors (sex, age, tissue, and cancer sub-type); under these analysis, all DE genes retained statistical significance (Supplementary Tables S29-S55). We used Hartigans' dip test to each gene to evaluate bi- or multi-modality in gene expression distributions which may imply presence of potential hidden variables that might affect expression of that gene (Additional File 1).

Differences in gene expression between populations are enriched for the inflammation/cytokines, endosomal development, and ROS metabolism network

GO terms related to the interrelated biological processes of inflammation/cytokines, endosomal development, and ROS metabolism are overrepresented among those genes that are DE between African Americans and European Americans (Supplementary Table S56). Similarly, Gene Set Enrichment Analysis (GSEA) of all of the 25 GTEx and TCGA sample-types shows KEGG pathways of immune- and inflammation-related processes are highly enriched (Supplementary Table S57-S58); the single most commonly-enriched pathway (found in 19 of the 25 sample-types) is "cytokine-cytokine receptor interaction" (Figure 1 and Additional File 2). The processes of oxidative metabolism/xenobiotic catabolism are enriched in 9 sample-types (Figure 1 and Additional File 2). Pooled GTEx data detects coordinated changes between African Americans and European Americans associated with four cytokine-related pathways and oxidative drug metabolism (Figure 1 and Supplementary Table S57).

The seven genes most highly and consistently DE between African Americans vs European Americans across all sample-types are: C-C Motif Chemokine Ligand CCL3L3, mitochondrial Glutathione-S-Transferase GSTM1; Nuclear Pore Complex Interacting Protein Family Member NPIP15, Coagulation Factor VIII Associated genes F8A3 and F8A2; FAM21B; and serine protease PRSS21. Of these, four are directly related to inflammation, endosomal development, and ROS metabolism. Each of these genes is also DE in multiple tissues and cancers.

A

tissue	infection/immune-response-related (# pathways enriched)	CK-CK interaction pathway	glutathione-oxidative-metabolic-processes (# pathways enriched)
kidney	4	yes	2
breast	6	no	3
esophagus	8	no	0
lung	8	no	2
uterus	8	yes	1
stomach	9	yes	3
thyroid	10	no	0
prostate	13	yes	0
colon	15	yes	0
liver	17	yes	0
pooledGTEx	10	yes	3

B

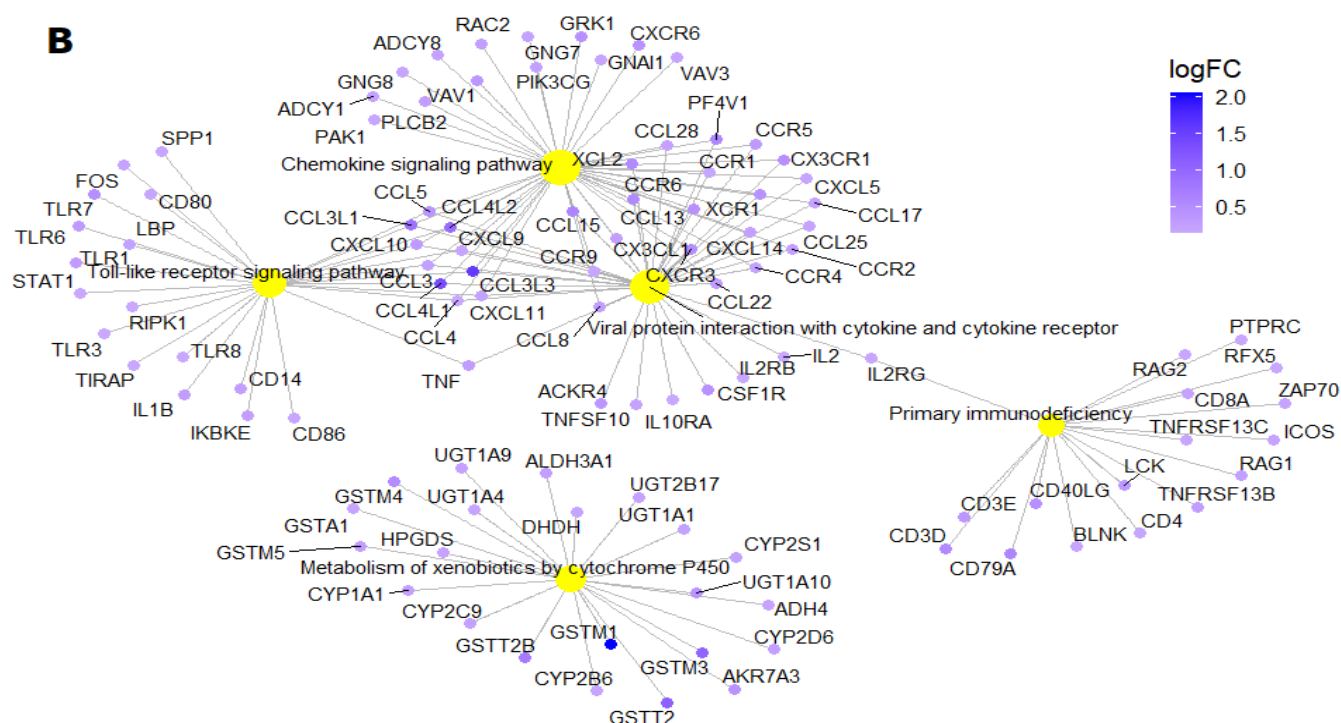


Figure 1. Gene Set Enrichment Analysis (GSEA) enrichment of KEGG pathways in African Americans compared to European Americans in pooled GTEx data. GSEA comprehensively analyses data for expression of all genes, rather than only the DE genes. **A.** The most common pathways enriched among upregulated genes in African Americans for tissue-types in GTEx. Complete list of enriched pathways and sample-types is in Additional File 2. CK-CK, cytokine-cytokine receptor interaction; glutathione-oxidative metabolism includes (oxidative) metabolism of xenobiotics. The full enrichment analysis for each sample-type is shown in Supplementary Table S57-S59. **B.** The five most highly enriched pathways among upregulated genes of pooled samples from all sample-types in GTEx are: Toll-like receptor signaling; chemokine signaling; primary immunodeficiency; viral protein interaction with cytokine and cytokine receptor; metabolism of xenobiotics by cytochrome P450.

Project	sample-type	#AA	#EA	#Upregulated	#Downregulated
GTEX	Breast	12	75	0	0
GTEX	Prostate	13	89	0	0
GTEX	Uterus	13	68	0	0
GTEX	Liver	15	97	0	0
GTEX	Stomach	29	159	4	6
GTEX	Colon	41	292	13	9
GTEX	Esophagus	80	564	19	11
GTEX	Thyroid	43	267	25	30
GTEX	Lung	39	269	45	20
TCGA	Lung squamous cell carcinoma (LUSC)	28	337	2	0
TCGA	Thyroid carcinoma (THCA)	25	292	3	3
TCGA	Lung adenocarcinoma (LUAD)	48	368	16	5
TCGA	Kidney renal papillary cell carcinoma (KIRP)	49	166	19	13
TCGA	Uterine Corpus Endometrial Carcinoma (UCEC)	54	70	28	5
TCGA	Colon adenocarcinoma (COAD)	54	188	30	21
TCGA	Kidney renal clear cell carcinoma (KIRC)	46	410	68	94
TCGA	Breast invasive carcinoma (BRCA)	142	674	83	164
GTEX	pooled GTEX samples	292	1,905	12	11
TCGA	pooled TCGA samples	497	3,238	13	21

Table 1. Number of DE genes in African Americans (AA) compared to European Americans (EA) in nine non-diseased tissue types and eight tumor types. Only sample-types with AA sample size 12 or greater are shown. Criteria for DE: >12 AA individuals sampled, >2-fold difference in expression based on Mann–Whitney U test (BH-corrected p-value < 0.05).

Cytokines, ROS and the storm

Multiple genes of the immune response are DE between African American and European American populations (Supplementary Table S2-S25). The small inducible chemokine, CCL3L3, is more highly expressed in African Americans by up to 7-fold in most diseased and non-diseased sample-types (Figure 2) (Supplementary Table S2-S25).

Several genes that mitigate oxidative stress are DE between African American and European American populations. In particular, GSTM1, a key enzyme involved in oxidative stress, is more highly expressed in African Americans than European Americans across multiple sample-types (Figure 2)).

F8As, endosome motility, and autophagy

Endocytosis and autophagy are intimately interrelated with COVID-19²². One little-studied player implicated in early endosome motility²³ is the F8A/HAP40 (HAP40) protein, encoded by three genes (F8A1, F8A2, and F8A3) in humans²⁴.

F8A1 is more highly expressed by about 2-fold in European Americans under almost every sample-type analyzed (Figure 3). Conversely, F8A2 and F8A3 are more highly expressed in African Americans. Expression of F8A2 in African Americans is up to 40-fold greater; expression of F8A3 is up to 6.6-fold greater. In LUSC, F8A2 and F8A3 are the only genes DE > 2-fold (Supplementary Table S7). F8A2 and F8A3 follow a similar trend, being more highly expressed in African Americans (Supplementary Table S2-S25 and **Figure 3 and Supplementary Figure 1**

Because of the vast differences in expression levels of the three HAP40-encoding genes between African Americans and European Americans, the paucity of literature on HAP40²⁵, and the relationships among F8A1, F8A2, and F8A3 genes, we further investigated the sequences, sequence variants, and the expression patterns of these genes.

The sequences of the HAP40-encoding proteins from F8A1, F8A2, and F8A3 are identical to each other in human reference genome GRCh38.p13 (https://www.ncbi.nlm.nih.gov/assembly/GCF_000001405.39). We searched for potential allele variants of HAP40 proteins encoded by F8A1, F8A2, and F8A3 in The Genome Aggregation Database (gnomAD)²⁶. gnomAD assigns individuals to populations, by clustering of genetic features. Our search identified only very rare sequence variants in the HAP40s encoded by F8A1, F8A2, or F8A3 (gnomAD v3). No structural variants were identified for HAP40 of F8A1 or F8A3; a duplication of 54 aa is, very rarely, present in F8A2 (gnomAD SVs v2.1).

To our knowledge, F8A1, F8A2 and F8A3 gene expression has never been compared. This may be because expression of F8A2 and F8A3 genes is relatively low in most European Americans, and European Americans are the predominant population studied. Furthermore, most RNA-Seq studies report expression of *only* F8A1 or F8A3 (and not F8A2), presumably aligning all

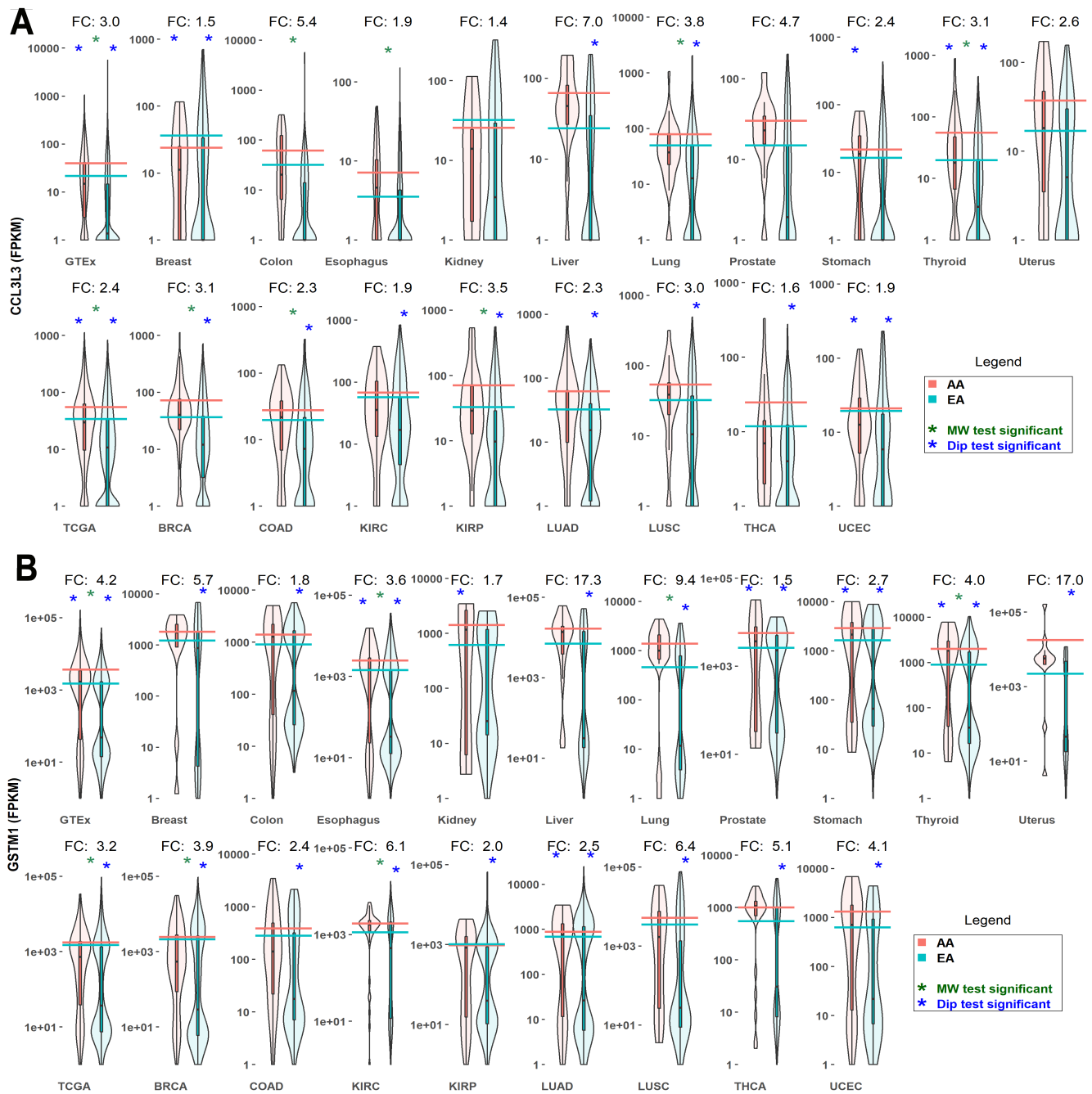


Figure 2. Upregulated expression of **CCL3L3** chemokine and mitochondrial glutathione-S-transferase **GSTM1** in African Americans compared to European Americans across multiple conditions. **A.** **CCL3L3** is more highly expressed in African Americans over a wide range of sample-types. **CCL3L3** binds to chemokine receptor proteins including CCR1, CCR3, and CCR5. **B.** **GSTM1** is more highly expressed in African Americans over a wide range of sample-types. **GSTM1** is a key player in metabolism of ROS and xenobiotics. (See Supplementary Tables S2-S28 for complete DE analysis). Violin plots summarize expression over each sample across the two populations. AA, African American; EA, European American. Horizontal lines represent mean log expression. *, MW test for DE significant (BH corrected p-value < 0.05). *, Hartigan's dip test, which assesses bimodal distribution potentially corresponding to hidden covariates. For a given gene and sample-type, a bimodal structure could imply presence of underlying hidden variables that affect expression of that gene, such as unreported sub-population structure or other environmental/genetic factors affecting gene expression in that multi-cellular organism. Expression distribution is also influenced by differences in population sizes (significant p-value < 0.05). FC, fold change AA/EA. GTEx and TCGA violin plots represent the pooled samples from each project. DE were computed within MetaOmGraph (MOG)¹⁸, in MOG's statistical analysis module; R scripts were executed interactively via MOG to generate the violin plots.

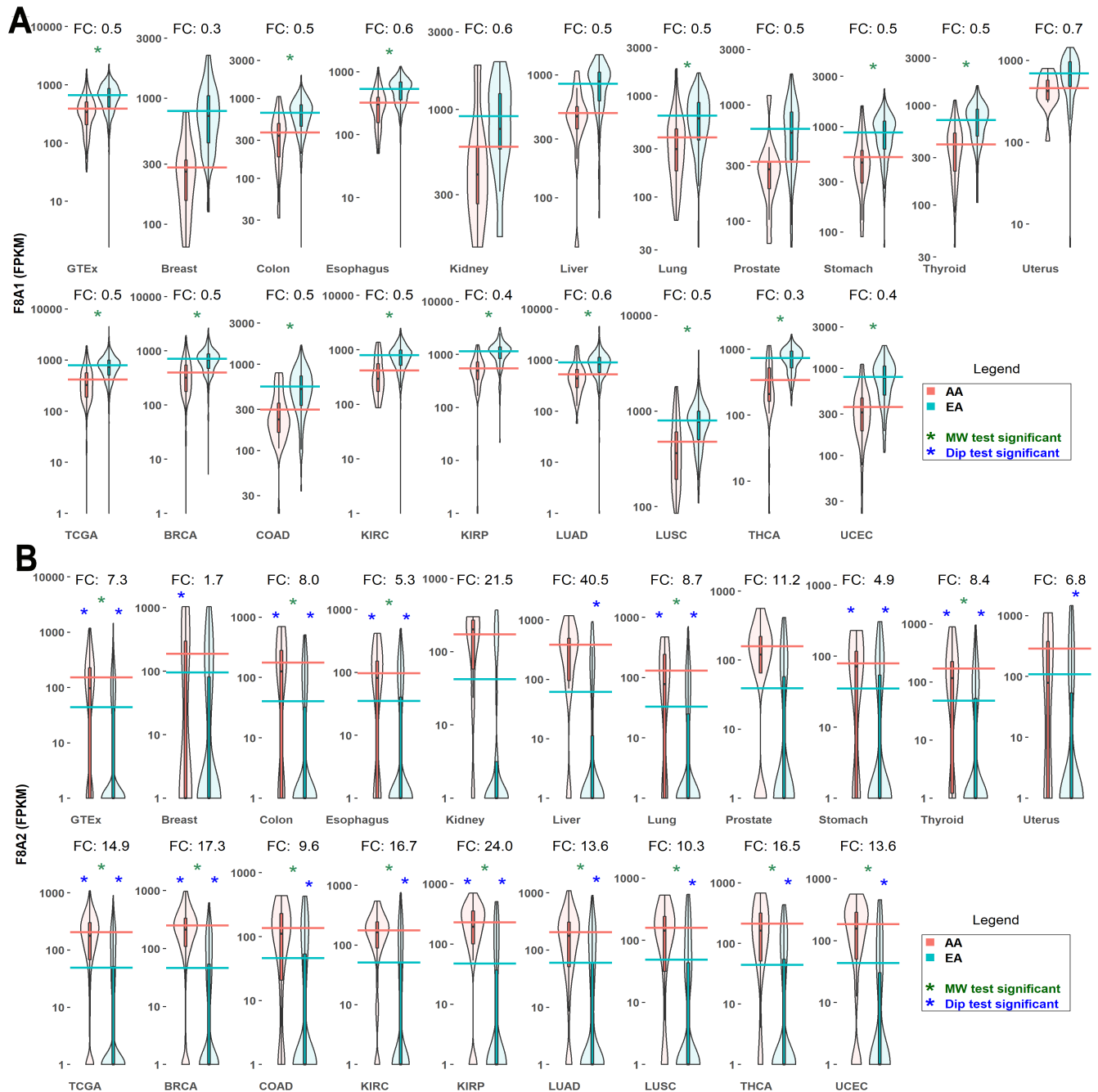


Figure 3. Differential expression of the HAP40 genes **F8A1** and **F8A2** in African Americans and European Americans across multiple sample-types. HAP40 is a key molecular component of Huntington's Disease, and shifts endosomal trafficking from the microtubules to actin fibers²³. **A.** **F8A1** expression is upregulated in European Americans. **B.** **F8A2** expression is upregulated in African Americans. Violin plots summarize expression over each sample across the two populations. AA, African American; EA, European American. Horizontal lines represent mean log expression. *, MW test for DE significant (BH corrected p-value < 0.05). *, Hartigan's dip test, which assesses bimodal distribution potentially corresponding to hidden covariates. For a given gene and sample-type, a bimodal structure could imply presence of underlying hidden variables that affect expression of that gene, such as unreported sub-population structure or other environmental/genetic factors affecting gene expression in that multi-cellular organism. Expression distribution is also influenced by differences in population sizes (significant p-value < 0.05). FC, fold change AA/EA. GTEx and TCGA violin plots represent the pooled samples from each project. DE were computed within MetaOmGraph (MOG)¹⁸, in MOG's statistical analysis module; R scripts were executed interactively via MOG to generate the violin plots. (See Supplementary Figure 2 for line plot comparison across individuals)

reads to one or the other gene.

We analyzed coexpression of the three F8A genes relative to the other 18,212 genes represented in the full TCGA-GTEX dataset using two statistical measures: Pearson's correlation and Mutual Information (MI). Although the three F8A genes are proximately located on the X chromosome, their expression patterns are *not* correlated. F8A2 and F8A3 have a Pearson's correlation of ($r = 0.40$), while F8A1 is negatively correlated with F8A2 and F8A3. Of all 18,212 genes represented in the data, the expression pattern of F8A1 is most *negatively* (anti-) correlated with that of F8A2 ($r = -0.45$) and F8A3, ($r = -0.24$) (Supplementary Table S60). F8A1 expression is not correlated with the F8 (Coagulation Factor FVIII) gene, although it resides with intron 22 of this gene. MI analysis indicates that F8A2 and F8A3 genes are more closely associated with F8A1 than with any other gene, consistent with the negative Pearson correlation (Supplementary Table S60).

Signatures of DE genes correspond to specific cell types in esophagus and lung

Because GTEx samples represent whole tissues, we sought to determine whether genes differentially expressed between African Americans and European Americans corresponded to distinct cell populations present in these tissues. This would provide information on cell-type representation across the two populations. To do this, we evaluated single cell datasets from two tissues highly relevant for SARS-CoV-2 infection: esophagus and lung^{27,28}.

Genes upregulated in African Americans in the esophagus map predominantly to two cell lineages, glandular epithelial cells of esophagus glands, and hematolymphoid lineage-associated dendritic cells (Figure 4). In proximal and distal airway cells of the lung, the signature of DE genes in African Americans versus European Americans corresponds to basal differentiating and proliferating keratinocytes (Figure 5).

Discussion

Human genetics contribute to the propensity and severity of diseases^{21,30–37}. Sometimes the contribution is straightforward; a single allele variation found in Ashkenazi Jews, causes the vast majority of Tay-Sachs disease³⁶. Sometimes it is more complex; for example, hypertension is more prevalent in African American than European American populations³⁷ in part due to detrimental APOL1 mutations that are more frequent in West African populations³¹. Despite the paucity of studies focused on Western African populations, the propensity and severity of other diseases among this population have been attributed to genetics^{21,31,38,39}.

Many COVID-19 deaths have been attributed to a cyclic over-excitement of the innate immune system^{1,7,16}. This process, often termed a cytokine storm, results in a massive production of cytokines, and the body attacking itself rather than specifically destroying the pathogen-containing cells^{1,7}. People with comorbidities, the elderly, and immunosuppressed individuals, may be at a greater risk for COVID-19 morbidity and mortality either because they may not respond to infection with a sufficient immune response⁴⁰ and/or because they may be more likely to develop a cytokine storm^{1,7}. Notably, many cytokines and other immunomodulatory molecules are DE between African Americans and European Americans in one or more sample-type, cytokine-related KEGG pathways are enriched for DE genes, and cell-type biomarkers indicate enrichment of DE genes in immune-related cell-types. Thus, predominant differences in gene expression, pathway enrichment, and cell-types between African Americans and European Americans are all implicated in biological processes that highly impact COVID morbidity and mortality.

CCL3L3, upregulated in African Americans under almost every diseased and non-diseased sample-type we tested, is also upregulated in COVID-19-diseased human bronchoalveolar lavage fluid⁴¹. The CCL3 protein, encoded by CCL3L3, is a member of the functionally-diverse C-C motif chemokine family and acts as ligand for CCR1, CCR3, and CCR5. CCL3 is a neutrophil chemotaxis protein, recruiting and activating granulocytes, and inhibiting HIV-1-infection⁴². CCL3L3 is upregulated in COVID-19, and neutrophils themselves are highly implicated in COVID-19 severity^{41,43}.

GSTM1, more highly expressed in African Americans compared to European Americans in almost every sample-type evaluated, is a key enzyme in mitochondrial ROS metabolism⁴⁴. Mitochondrially-generated ROS induce expression of proinflammatory cytokines and chemokines, and are considered to play a key role in modulating innate immune responses against RNA viruses⁴⁴ including SARS-COV-2⁴⁵. Higher expression of GSTM1, could lead to increased mitochondrial ROS, which might ultimately trigger inflammation and a cytokine storm⁴⁴. Alternatively, higher GSTM1 expression might cause ROS to be metabolized rapidly, and prevented ROS from initiating a sufficient immune response. GSTM1 has a second critical function— in metabolism of xenobiotics, including many toxins and pharmaceuticals⁴⁴. Thus, if GSTM1 is highly expressed, pharmaceuticals may be rapidly metabolized and rendered inactive.

The most dramatic differences in genes expression in African Americans compared to European Americans are associated with the highly-conserved but little-studied F8A genes, which each encode HAP40. F8A1 is upregulated in European Americans, while F8A2 and F8A3 are upregulated in African Americans. HAP40 function has been researched mostly in the context of F8A1 and the critical role of that gene in slowing early endosome mobility in Huntington's disease²⁴. In Huntington's, HAP40 forms a bridge between the huntingtin protein and the regulatory small guanosine triphosphatase, RAB5; formation of this

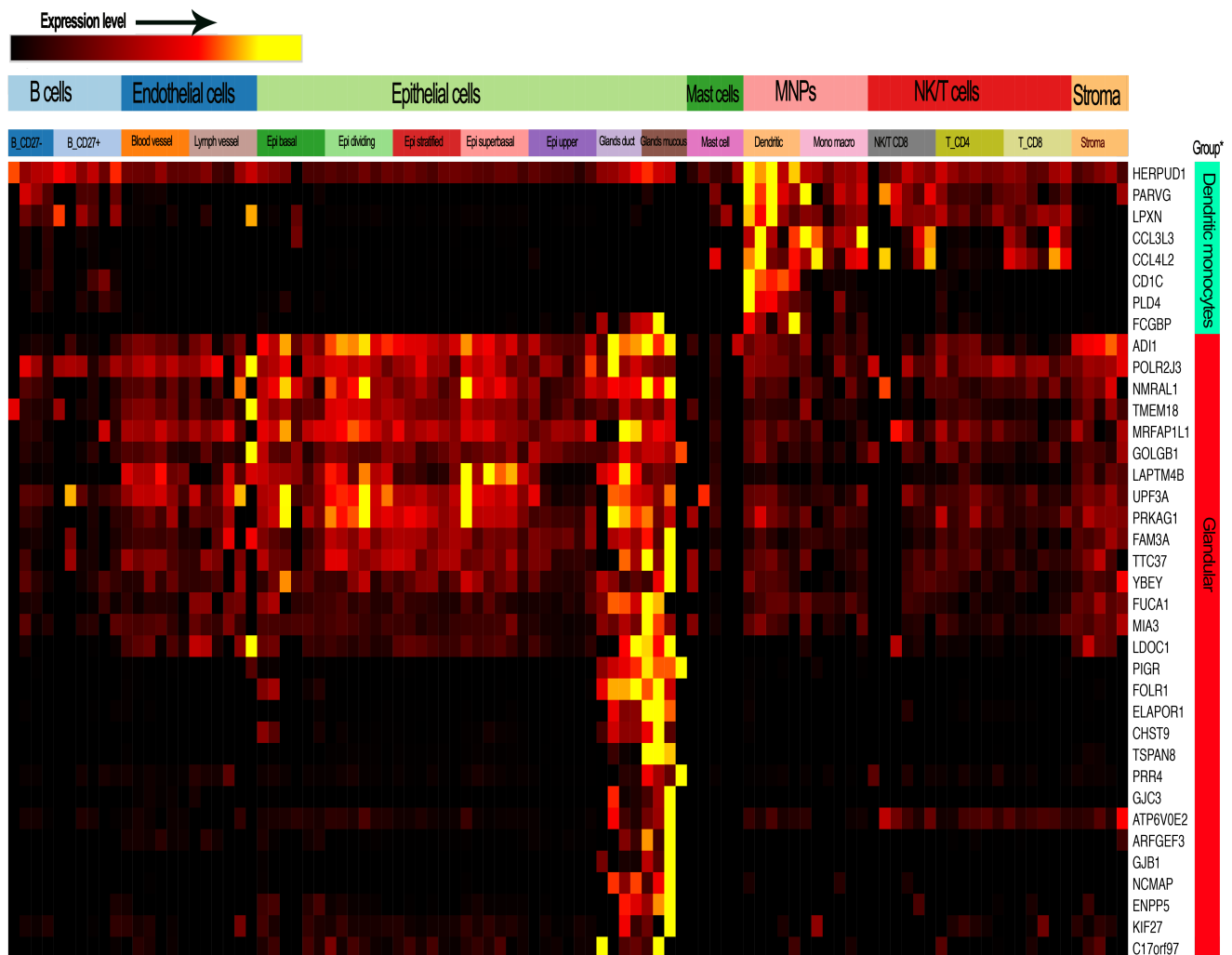


Figure 4. Esophageal gene signatures vary in African Americans and European Americans in a cell-specific manner. Gene signature upregulated in African American versus European American esophagus maps to two cell lineages with prominent presence in the human cell atlas (<https://data.humancellatlas.org/>) esophageal dataset. One significant fraction of the African American-upregulated gene signature maps to glandular mucous epithelial cells of esophageal glands (genes marked by red, far right bar). Expression of several of the genes upregulated in African Americans is highly restricted to the mucous epithelial cells (TSAPN8, PRR4, ELAPOR1), whereas FOLR1, for example, is more highly expressed in the ductal epithelial cells of the gland. A second, smaller, signature corresponds to hematolymphoid/myeloid lineage dendritic cells, as shown by CDC1C, PLD4, HERPUD1, and LPXN (genes marked by green, far right bar). In addition to a number of genes that are most strongly expressed by those cell types, there are several genes that are essentially exclusively expressed by those cells. ToppCell-constructed gene modules (<http://toppcell.cchmc.org>) for each of the cell types reported to be present in the large scRNA-Seq dataset from esophagus²⁹.



Figure 5. Lung gene signatures upregulated in African Americans versus European Americans map to proximal airway keratinocytic epithelial lineage, and to mesenchymal mesothelial and neuroendocrine cells. Marker genes for keratinocytes (genes marked by yellow, far right bar); ciliated epithelial cells (genes marked by turquoise, far right bar); mesothelial mesenchymal cells (genes marked by red, far right bar); and neuroendocrine mesenchymal cells (mesenchymal). Note that the keratinocytic proximal basal epithelial cell is the cell subtype with the highest expression of ACE2 receptor, a major target of COVID-19 (ACE2 marked by black on bar at right). ToppCell-constructed gene modules (<http://toppcell.cchmc.org>) for each of the cell types reported to be present in the large scRNA-Seq dataset from lung²⁸.

complex reduces endosomal motility by shifting endosomal trafficking from the microtubule to the actin cytoskeleton²³. F8A1 overexpression in striatal neuron cell lines from mice resulted in increased ROS and mitochondrial dysfunction⁴⁶. Knockouts of F8A1 in human HeLa and HEK293 cells yield altered/reduced autophagy and shorter life spans⁴⁶. Knockouts of the single F8A gene in *Drosophila* similarly show reduced activity, altered/reduced autophagy, and shorter lifespan⁴⁷.

F8A1 expression is increased under several conditions, including Huntington's disease⁴⁸, presence of a SNP variant for type 1 diabetes risk⁴⁹, cytotrophoblast-enriched placental tissues in women with severe preeclampsia⁵⁰, and mesenchymal bone marrow cells as women age⁵¹. Its potential role in the latter conditions has not been investigated.

Altered endosome motility would play an important but complex role in infection and the innate immune response, and might either promote or hinder the battle between SARS-CoV-2 and its human host^{22,52}. Coronaviruses including SARS-CoV-2 mainly enter host cells via binding to the ACE2 receptor followed by endocytosis^{7,53}. Nascent early endosomes are moved along the microtubule cytoskeleton, fusing with other vesicles; varied molecules can be incorporated into the endosomal membrane or its interior^{22,52}. This regulated development enables diverse fates. For example, in the context of SARS-CoV-2, endosomes might release viral RNA or particles; they might merge with lysosomes and digest their viral cargo; or they might fuse with autophagosomes (autophagy) and subsequently with lysosomes that digest the cargo^{22,52}. SARS-CoV-2 might reprogram cellular metabolism to suppress autophagy and promote viral replication⁵⁴; conversely, the cell might modify autophagy machinery to decorate viral invaders with ubiquitin for eventual destruction, activate the immune system by displaying parts of the virus, or catabolize excess pro-cytokines. Autophagy might induce cytokine signaling, which could promote protective immune response or engender a destructive storm of cytokines, inflammation and tissue damage²². Because of its function in early endosome motility, HAP40 has implications as a potential molecular target in therapy of endosomal and autophagy-related disorders such as COVID-19.

Our analyses using single cell reference data indicate several cell type-specific associations of the signatures of DE genes in African Americans versus European Americans. One model by which this might occur is that individuals of one population tend to have different proportions of a given cell type or histological structure. An alternative model is that individuals of one population might tend to maintain some of their cell types in a state of relatively higher activation. Either explanation would lead bulk RNA-Seq analyses, such as sample-types from GTEx or TCGA, to demonstrate elevated expression of those transcripts in that population.

Although at a population level, major differences exist in expression of immunity-related genes and cell-type-specific associations between African Americans and European Americans, when considered on an individual basis, gene expression differences are more complex. Individuals within a population may exhibit all, none, or some portion of these differences. That some genes show bimodal expression distribution in some sample-types African American and/or European American populations further emphasizes this variation.

Thus, the significance of these patterns and their relationship to differential susceptibility or risk of severity from COVID-19 infection must be considered from nuanced perspectives. Importantly, it may be that only a fraction of the signature and a fraction of the individuals in a population are at elevated risk of more severe disease. In addition, different mechanisms of risk may be operative within different individuals within a population. For example, elevated abundance or activity of cells that are the target of COVID-19 (e.g., ACE2-positive basal keratinocytes) could lead to a greater infection burst during initial phases with a larger number of virions being released systemically. If, as it appears from the alignment of the DE genes in African American compared to European Americans to the lung single cell data, this is the case for African American-individuals, then they might be more readily taken over by infecting SARS-Cov2 virions.

The differential expression of genes implicated in COVID-19 morbidity and mortality between African Americans and European Americans reported herein emphasizes the importance of integrating gene expression data into the factors considered in studying this pandemic at a population level. Further, RNA-Seq data has been shown useful in clinical practice for pediatric cancers⁵⁵, and this practice could be extended to other diseases. We hypothesize, in concurrence with^{17,56}, that processes of disease- and stress-related genes are overrepresented among the DE genes of African American and European American populations in part because the ancestral selection pressure due to disease and stresses (such as temperature and toxins) was very strong, with very different complements of pathogens and stresses in the regions where these populations lived. To survive, humans living in Europe and those living in Western Africa would have had to evolve the ability to resist the diverse prevalent local pathogens and stresses. Other differences would be due to a difference in socioeconomic environment⁵⁷.

The utility of expression data is tremendous, but it is reliant on adequate *representation of cohorts* and on sufficient *metadata*. For example, ethnic bias and practical factors (such as subject availability) often result in insufficient numbers of subjects from many populations being represented in medical studies^{58,59}; this lack of representation impedes the development of precision prognosis and therapy based on genetics^{34,58}. Here, we were limited to comparison of differences between gene expression in African American and European American populations because even in the large GTEx and TCGA studies, sample sizes for the other three major population groups (Asian, Native American, and Pacific Islanders) were generally too low for robust statistical assessment (Supplementary Table S1).

In addition, even if sample sizes for race are sufficient, information on the ancestry of each individual sampled is needed. Self-reported metadata on race is often not publicly available for individual samples. However, methods of assigning ancestry to individuals sampled for RNA-Seq are being developed and applied^{60,61}

A major dichotomy exists between socio-economic and genomic investigations. Among the vast body of human RNA-Seq data deposited, not only are metadata on the ancestry of the sampled individuals often unavailable, but socio-economic metadata (postal code, education, income, occupation) are almost never present. Thus, apart from the pioneering sociogenomics research of^{57,57,62} and the studies of^{56,63,64}, socioeconomic information are rarely considered in 'omics analyses. Indeed, because of the scant metadata on socioeconomic determinants it is not even possible to determine possible skewness of representation of socioeconomic groups among the individuals sampled; thus, socioeconomic factors represent high-impact complex hidden covariates that would be challenging to model. Conversely, sociological studies rarely incorporate 'omics information. For example, the U.S.-based Robert Wood Johnson Foundation (<https://www.rwjf.org/en/library/interactives/whereyouliveaffectshowlongyoulive.html>) cites research that "your zip code can be more important than your genetic code" for your health; however, the analyses were done without actually evaluating genetic codes.

Because socioeconomic data was absent in this study, we were unable to distinguish genetic effects from socioeconomic causes. Despite having metadata on ancestry, we could not resolve the component due to genetics, from that to socioeconomic factors, and were limited to reporting population-based differences in gene expression (rather than ancestry-based differences). Privacy concerns need to be carefully balanced against the very real health benefits that can be gained from metadata access. Without routine inclusion and availability of diverse metadata for human 'omics samples, data mining is hampered, and important information is lost.

In summary, multiple genes implicated in COVID-19 immunity and inflammation are DE across African American and European American populations. This differential expression is evident despite the fact that race is self-reported in the metadata, and many Americans are racially admixed²¹. By highlighting the wide-ranging differences in expression of genes implicated in the morbidity and mortality of COVID-19 across populations, and by revealing apparent cell-type differences between populations, we provide a baseline for future study and emphasize the importance of harvesting these types of information for medicine. Such research will establish prognostic signatures with vast implications for precision treatment of diseases such as COVID-19.

Online Methods

Datasets

GTEx provides data representing "non-diseased" samples from diverse tissues. Non-diseased refers to the tissue itself, however, in some cases the individual sampled was postmortem and the causes of death are varied. TCGA project is the largest project available on different diseased samples (tumors) of multiple tissue origins. Both projects have metadata on the (self-reported) races of the individuals who contributed samples. These two projects provide a unique opportunity to evaluate differences in gene expression across populations in multiple sample-types that vary by tissue and disease status. Tissues and cancers were selected for downstream analysis based largely on having sufficient numbers of individuals from each ancestry. We refer to those self-reporting as "Black or African American" as "African Americans" and "White" as "European Americans".

The data files and the precompiled MOG project, *MOG_HumanCancerRNASeqProject*, were downloaded from http://metnetweb.gdcb.iastate.edu/MetNet_MetaOmGraph.htm¹⁸. This project uses batch-corrected and processed data to enable comparison across samples¹⁹. *MOG_HumanCancerRNASeqProject* contains expression values for 18,212 genes, 30 fields of metadata detailing each gene, across 7,142 samples representing 14 different cancer types and associated non-tumor tissues (TCGA and GTEX samples) integrated with 23 fields of metadata describing each study and sample¹⁸.

Statistical and correlation analyses

The MOG tool was used to interactively explore, visualize and perform differential expression and correlation analysis of genes.

The Mann-Whitney (MW) test was used to identify DE genes between two groups; we chose this non-parametric analysis as it makes no assumptions about the data distribution. We define a gene as DE 2-fold or more between two groups if it meets each of the following criteria:

1. Estimated fold-change in expression of 2-fold or more (log fold change, $|\log FC| \geq 1$), where $\log FC$ is calculated as in `limma`⁶⁵.)
2. Mann–Whitney U test is significant between the two groups (Benjamini-Hochberg (BH) corrected p-value < 0.05)

Pearson correlation values and Mutual Information values were computed after data was \log_2 transformed within MOG, in MOG's statistical analysis module. R scripts were written to create the violin plots; these scripts were executed interactively via MOG.

Covariate evaluation

To check for potential sampling differences between populations that might confound the analysis, we fit linear models using limma⁶⁵ in R, to adjust for biologically relevant, potential confounding factors of race, gender, tissue/tumor type, age and cancer subtypes. (Supplementary Table S29-S55). Because ratios of cancer sub-types may differ between races (as reported for breast cancer in premenopausal African American women), we evaluated the RNA-Seq data from African Americans and European Americans in BRCA samples for potential confounding effects due to different ratios of four breast cancer subtypes: basal-like (BAS), human epidermal growth factor receptor-2 positive/estrogen receptor negative (Her2), luminal A (LumA), and luminal B (LumB); all genes DE with an >2-fold change in MW analysis retained statistical significance in limma analysis of BRCA data, although the fold-change level varied (Supplementary Table S41).

To assess whether a given distribution shows bi- or multi-modality we applied the Hartigans' dip (Dip) test, using the R package diptest (<https://cran.r-project.org/package=diptest>) (Additional File 1).

Gene expression enrichment

Overrepresentation of biological processes and other functional analysis was assessed at <https://toppgene.cchmc.org/>. Geneset enrichment analyses (GSEA) were performed using the clusterProfiler library in R⁶⁶.

Cell-type analysis

African American-vs-European American gene signatures to cell types and compartments were mapped using cell type specific gene modules derived from a series of single cell gene expression datasets for esophagus²⁹ and lung²⁸ hosted in ToppCell (<http://toppcell.cchmc.org/>) using the ToppGene tool (<http://toppgene.cchmc.org/>). Heat map visualization of genes differentially-expressed by African Americans versus European Americans in each cell type module in the selected tissues was done using Morpheus (<https://software.broadinstitute.org/morpheus/>) using ToppCell's "super binned" gene expression for each cell type within each single cell dataset.

Data availability

We subscribe to an open data model (<https://www.go-fair.org/fair-principles/>). MOG is free and open source software published under the MIT License. MOG software, user guide, and the *MOG_HumanCancerRNASeqProject* project datasets and metadata described in this article are freely downloadable from http://metnetweb.gdcdb.iastate.edu/MetNet_MetaOmGraph.htm. MOG's source code is available at <https://github.com/urmi-21/MetaOmGraph/>. Detailed information and code on how to reproduce the results, along with Additional files, are available at <https://github.com/urmi-21/COVID-DEA>.

Supplementary data

Supplementary data are available at <https://github.com/urmi-21/COVID-DEA>.

Funding

This work is funded in part by the National Science Foundation award IOS 1546858, "Orphan Genes: An Untapped Genetic Reservoir of Novel Traits" and by the Center for Metabolic Biology, Iowa State University. This work used the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by National Science Foundation grant number ACI-1548562. In particular, it used the Bridges HPC environment through allocations TG-MCB190098 and TG-MCB200123 awarded from XSEDE and the HPC Consortium.

Acknowledgements

We are grateful to Mashette Syrkin-Nikolau, Diane Bassham, Karin Dorman and Judy Syrkin-Nikolau for valuable discussions and comments on the manuscript. We thank Afshin Beheshti and Shawn Brown for providing leadership and encouragement in combating COVID. We are grateful for the collaborate spirit of the COVID-19 International Research Team (COVIRT) (<https://covirt19.org/>) and the International COVID-19 Knowledge Exchange group, and the many helpful interactions we have with researchers in these and other consortia.

Author contributions statement

US conceived the study, analyzed the data, and wrote the paper.
KH performed GSEA analysis and commented on the paper.

BJA developed and performed the cell-specific profile analysis, wrote the parts of the paper he contributed to, and commented on the paper.

ESW conceived the study, analyzed the data, and wrote the paper.

Conflict of interest statement.

None declared.

References

1. Jeyanathan, M. *et al.* Immunological considerations for covid-19 vaccine strategies. *Nat. Rev. Immunol.* 1–18 (2020).
2. Bansal, M. Cardiovascular disease and covid-19. *Diabetes & Metab. Syndr. Clin. Res. & Rev.* (2020).
3. Varga, Z. *et al.* Endothelial cell infection and endotheliitis in covid-19. *The Lancet* **395**, 1417–1418 (2020).
4. Snell, J. Sars-cov-2 infection and its association with thrombosis and ischemic stroke: A review covid-19, thrombosis, and ischemic stroke. *The Am. J. Emerg. Medicine* (2020).
5. Frithiof, R., Bergqvist, A., Järhult, J. D., Lipcsey, M. & Hultström, M. Presence of sars-cov-2 in urine is rare and not associated with acute kidney injury in critically ill covid-19 patients. *Critical Care* **24**, 1–3 (2020).
6. Jarrahi, A. *et al.* Neurological consequences of covid-19: what have we learned and where do we go from here? *J. Neuroinflammation* **17**, 1–12 (2020).
7. Tay, M. Z., Poh, C. M., Rénia, L., MacAry, P. A. & Ng, L. F. The trinity of covid-19: immunity, inflammation and intervention. *Nat. Rev. Immunol.* 1–12 (2020).
8. Teuwen, L.-A., Geldhof, V., Pasut, A. & Carmeliet, P. Covid-19: the vasculature unleashed. *Nat. Rev. Immunol.* 1–3 (2020).
9. Del Valle, D. M. *et al.* An inflammatory cytokine signature predicts covid-19 severity and survival. *Nat. medicine* 1–8 (2020).
10. Millett, G. A. *et al.* Assessing differential impacts of covid-19 on black communities. *Annals Epidemiol.* (2020).
11. Battagello, D. S. *et al.* Unpuzzling covid-19: tissue-related signaling pathways associated with sars-cov-2 infection and transmission. *Clin. Sci.* **134**, 2137–2160 (2020).
12. Williams, F. M. *et al.* Self-reported symptoms of covid-19 including symptoms most predictive of sars-cov-2 infection, are heritable. *medRxiv* (2020).
13. Ellinghaus, D. *et al.* Genomewide association study of severe covid-19 with respiratory failure. *New Engl. J. Medicine* (2020).
14. Warren, R. L. & Birol, I. Hla predictions from the bronchoalveolar lavage fluid samples of five patients at the early stage of the wuhan seafood market covid-19 outbreak. *arXiv preprint arXiv:2004.07108* (2020).
15. Devaux, C. A., Rolain, J.-M. & Raoult, D. Ace2 receptor polymorphism: Susceptibility to sars-cov-2, hypertension, multi-organ failure, and covid-19 disease outcome. *J. Microbiol. Immunol. Infect.* (2020).
16. Woo, Y.-L. *et al.* A genetic predisposition for cytokine storm in life-threatening covid-19 infection. (2020).
17. Nédélec, Y. *et al.* Genetic ancestry and natural selection drive population differences in immune responses to pathogens. *Cell* **167**, 657–669 (2016).
18. Singh, U., Hur, M., Dorman, K. & Wurtele, E. S. Metaomgraph: a workbench for interactive exploratory data analysis of large expression datasets. *Nucleic Acids Res.* **48**, e23–e23 (2020).
19. Wang, Q. *et al.* Unifying cancer and normal rna sequencing data from different sources. *Sci. data* **5**, 180061 (2018).
20. Baharian, S. *et al.* The great migration and african-american genomic diversity. *PLoS genetics* **12** (2016).
21. Zhong, Y. *et al.* Discovery of novel hepatocyte eqtls in african americans. *PLoS genetics* **16**, e1008662 (2020).
22. Carmona-Gutierrez, D. *et al.* Digesting the crisis: autophagy and coronaviruses. *Microb. Cell* **7**, 119 (2020).
23. Pal, A., Severin, F., Lommer, B., Shevchenko, A. & Zerial, M. Huntingtin–hap40 complex is a novel rab5 effector that regulates early endosome motility and is up-regulated in huntington’s disease. *The J. cell biology* **172**, 605–618 (2006).
24. Perez-Riba, A. & Itzhaki, L. S. The tetratricopeptide-repeat motif is a versatile platform that enables diverse modes of molecular recognition. *Curr. opinion structural biology* **54**, 43–49 (2019).

25. Furr-Stimming, E., Shiyu, X., Ye, X., Zhang, S. *et al.* Hap40 is a conserved partner and regulator of huntingtin and a pathogenic modifier of huntington's disease (2817) (2020).
26. Karczewski, K. J. *et al.* The mutational constraint spectrum quantified from variation in 141,456 humans. *bioRxiv* 531210 (2020).
27. Jimenez, L. *et al.* The influence of ph on sars-cov-2 infection and covid-19 severity. *medRxiv* (2020).
28. Travaglini, K. J. *et al.* A molecular cell atlas of the human lung from single cell rna sequencing. *BioRxiv* 742320 (2020).
29. Madisson, E. *et al.* scrna-seq assessment of the human lung, spleen, and esophagus tissue stability after cold preservation. *Genome biology* **21**, 1–16 (2020).
30. Amorim, C. E. G. *et al.* The population genetics of human disease: The case of recessive, lethal mutations. *PLoS genetics* **13** (2017).
31. Kruzel-Davila, E., Wasser, W. G. & Skorecki, K. Apol1 nephropathy: a population genetics and evolutionary medicine detective story. In *Seminars in nephrology*, vol. 37, 490–507 (Elsevier, 2017).
32. Paulucci, D. J. *et al.* Genomic differences between black and white patients implicate a distinct immune response to papillary renal cell carcinoma. *Oncotarget* **8**, 5196 (2017).
33. Wu, M. *et al.* Race influences survival in glioblastoma patients with kps 80 and associates with genetic markers of retinoic acid metabolism. *J. neuro-oncology* **142**, 375–384 (2019).
34. Barrow, M. A. *et al.* A functional role for the cancer disparity-linked genes, cry β 2 and cry β 2p1, in the promotion of breast cancer. *Breast Cancer Res.* **21**, 1–13 (2019).
35. Chi, C. *et al.* Admixture mapping reveals evidence of differential multiple sclerosis risk by genetic ancestry. *PLoS genetics* **15**, e1007808 (2019).
36. O'Brien, J. S. *et al.* Tay-sachs disease: prenatal diagnosis. *Science* **172**, 61–64 (1971).
37. Burt, V. L. *et al.* Prevalence of hypertension in the us adult population: results from the third national health and nutrition examination survey, 1988-1991. *Hypertension* **25**, 305–313 (1995).
38. Need, A. C. & Goldstein, D. B. Next generation disparities in human genomics: concerns and remedies. *Trends Genet.* **25**, 489–494 (2009).
39. Backer, A. Why covid-19 may be disproportionately killing african americans: Black overrepresentation among covid-19 mortality increases with lower irradiance, where ethnicity is more predictive of covid-19 infection and mortality than median income. *Where Ethn. Is More Predict. COVID-19 Infect. Mortal. Than Median Income (April 8, 2020)* (2020).
40. Ahmadpoor, P. & Rostaing, L. Why the immune system fails to mount an adaptive immune response to a covid-19 infection. *Transpl. Int.* (2020).
41. Didangelos, A. Covid-19 hyperinflammation: What about neutrophils? *mSphere* **5** (2020).
42. Struyf, S. *et al.* Diverging binding capacities of natural ld78 β isoforms of macrophage inflammatory protein-1 α to the cc chemokine receptors 1, 3 and 5 affect their anti-hiv-1 activity and chemotactic potencies for neutrophils and eosinophils. *Eur. journal immunology* **31**, 2170–2178 (2001).
43. Veras, F. P. *et al.* Sars-cov-2 triggered neutrophil extracellular traps (nets) mediate covid-19 pathology. *medRxiv* (2020).
44. Sies, H. & Jones, D. P. Reactive oxygen species (ros) as pleiotropic physiological signalling agents. *Nat. Rev. Mol. Cell Biol.* 1–21 (2020).
45. Schönrich, G., Raftery, M. J. & Samstag, Y. Devilishly radical network in covid-19: Oxidative stress, neutrophil extracellular traps (nets), and t cell suppression. *Adv. Biol. Regul.* 100741 (2020).
46. Huang, Z.-N., Chung, H. M., Fang, S.-C. & Her, L.-S. Adhesion regulating molecule 1 mediates hap40 overexpression-induced mitochondrial defects. *Int. journal biological sciences* **13**, 1420 (2017).
47. Xu, S. *et al.* Hap40 is a conserved central regulator of huntingtin and a specific modulator of mutant huntingtin toxicity. *bioRxiv* (2020).
48. Peters, M. F. & Ross, C. A. Isolation of a 40-kda huntingtin-associated protein. *J. Biol. Chem.* **276**, 3188–3194 (2001).
49. Brumpton, B. M. & Ferreira, M. A. Multivariate eqtl mapping uncovers functional variation on the x-chromosome associated with complex disease traits. *Hum. genetics* **135**, 827–839 (2016).
50. Gormley, M. *et al.* Preeclampsia: novel insights from global rna profiling of trophoblast subpopulations. *Am. journal obstetrics gynecology* **217**, 200–e1 (2017).

51. Roforth, M. M. *et al.* Global transcriptional profiling using rna sequencing and dna methylation patterns in highly enriched mesenchymal cells from young versus elderly women. *Bone* **76**, 49–57 (2015).
52. Tao, S. & Drexler, I. Targeting autophagy in innate immune cells: Angel or demon during infection and vaccination? *Front. Immunol.* **11**, 460 (2020).
53. Chowdhury, R. & Maranas, C. D. Biophysical characterization of the sars-cov2 spike protein binding with the ace2 receptor explains increased covid-19 pathogenesis. *bioRxiv* (2020).
54. Gassen, N. C. *et al.* Analysis of sars-cov-2-controlled autophagy reveals spermidine, mk-2206, and niclosamide as putative antiviral therapeutics. *bioRxiv* (2020).
55. Vaske, O. M. *et al.* Comparative tumor rna sequencing analysis for difficult-to-treat pediatric and young adult patients with cancer. *JAMA network open* **2**, e1913968–e1913968 (2019).
56. Quintana-Murci, L. Human immunology through the lens of evolutionary genetics. *Cell* **177**, 184–199 (2019).
57. Cole, S. W. Human social genomics. *PLoS genetics* **10** (2014).
58. Friedman, P. N. *et al.* The acco u nt consortium: A model for the discovery, translation, and implementation of precision medicine in african americans. *Clin. translational science* **12**, 209–217 (2019).
59. McGuire, A. L. *et al.* The road ahead in genetics and genomics. *Nat. Rev. Genet.* 1–16 (2020).
60. Yuan, J. *et al.* Integrated analysis of genetic ancestry and genomic alterations across cancers. *Cancer cell* **34**, 549–560 (2018).
61. Barral-Arca, R., Pardo-Seco, J., Bello, X., Martinon-Torres, F. & Salas, A. Ancestry patterns inferred from massive rna-seq data. *RNA* **25**, 857–868 (2019).
62. Dieckmann, L., Cole, S. & Kumsta, R. Stress genomics revisited: gene co-expression analysis identifies molecular signatures associated with childhood adversity. *Transl. psychiatry* **10**, 1–11 (2020).
63. Favé, M.-J. *et al.* Gene-by-environment interactions in urban populations modulate risk phenotypes. *Nat. communications* **9**, 1–12 (2018).
64. Hooten, N. N. & Evans, M. K. Age and poverty status alter the coding and noncoding transcriptome. *Aging (Albany NY)* **11**, 1189 (2019).
65. Ritchie, M. E. *et al.* limma powers differential expression analyses for rna-sequencing and microarray studies. *Nucleic acids research* **43**, e47–e47 (2015).
66. Yu, G., Wang, L.-G., Han, Y. & He, Q.-Y. clusterprofiler: an r package for comparing biological themes among gene clusters. *Omics: a journal integrative biology* **16**, 284–287 (2012).